

# **Internship : co-clustering for functional data**

Laboratoire ERIC Université Lyon 2, in collaboration with EDF & Thalès,

## **1. Context**

After the installation of 300.000 smart meters "Linky" between 2009 and 2011 in the area of Lyon and Tours, the authorities have decided to generalize these meters throughout the territory. By 2021, 35 million meters should be replaced in french households by Linky meters, allowing electricity operators to record electricity consumption. For an operator like EDF with 27 millions of residential dwellings, these new smart meters represent a great opportunity to gather customer's consumption data and therefore to improve client knowledge. Indeed, so far, customer data were recorded only every six months, while with the smart meter, the data can be taken every half hour, which means 17472 measures for each of the 27 million clients in a year. Nevertheless, this data flood may also be a drawback since they represent a mass of data to store and manage. To this end, it will be necessary to build « summaries » of these data, and one of the way to achieve that is to cluster the data. However, because of the nature of the data, which are time series for each customer, the interest in the simultaneous clustering of customers and time increases considerably.

## **2. Subject**

The goal of the internship is to contribute to develop a co-clustering approach for functional data analysis. A first approach has already been developed, which relies on the clustering model for functional data developed in Bouveyron & Jacques 2011. Several works have to be carry out in order to improve this earlier approach.

The main missions are :

- to develop a strategy for taking into account the missing data,
- to adapt the model in order to be possible to use different types of basis of functions, as wavelets for instance,

A publication presenting both the model improvements and the package will be written during the internship.

The intern should have high skills in statistics and in R programming. A knowledge of model-based approach for clustering would be appreciated.

## **3. References**

C. Bouveyron, E. Côme and J. Jacques (2015), The discriminative functional mixture model for the analysis of bike sharing systems, *Annals of Applied Statistics*, 9[4], 1726-1760.

C. Bouveyron and J. Jacques (2011), Model-based Clustering of Time Series in Group-specific Functional Subspaces, *Advances in Data Analysis and Classification*, 5[4], 281-300.

## **4. Internship conditions**

location : the intern will join the Data Mining & Decision team of the ERIC lab., which is composed of 12 permanent researchers in statistics and computer science.

duration: 6 months, starting in March 2017

salary: approx. 550€ / month

contact: [julien.jacques@univ-lyon2.fr](mailto:julien.jacques@univ-lyon2.fr)